DOCUMENT RESUME

ED 113 631 95 CG 010 133

AUTHOR Cline, Victor B.
TITLE Process and Accuracy in Person Perception. Final Report.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
BUREAU NO BR-8-0769
PUB DATE [73]
GRANT OEG-9-540769-4277(010)
NOTE 48p.

EDRS PRICE MF-$0.76 HC-$1.95 Plus Postage
DESCRIPTORS *Behavior Patterns; *Motivation; Nonverbal Communication; *Perception; Psychological Testing; Research Projects; *Stereotypes; Stimulus Behavior; *Verbal Communication; Visual Perception

ABSTRACT

This final report attempts to summarize the major findings of three and one-half years of research in the area of understanding and accurately judging others. The author's research on person perception focuses on the use of sound motion pictures of interviews as the primary stimulus and vehicle for the research. Subjects used for these studies were usually college students from psychology classes. There are three major sections or topics for this research: (1) assessing the validity of verbal and non-verbal cues in accurately judging others; (2) effects of money, anticipation and competition upon judging scores and self-reports of motivational arousal; and (3) social stereotyping and its relationship to accuracy in person perception. Results indicate that accurately perceiving and judging another is a complex task affected by many variables, and the difficulty in decoding them depends on the transparency or opaqueness of the person being judged. In addition, according to the authors, people can be trained to be better judges of others. (GMF)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED113631

CG

FINAL REPORT

To: Office of Education, Bureau of Research, Basic Studies Branch, Division of Comprehensive and Vocational Education Research
Title: Process and Accuracy in Person Perception
Grant Number: OEG-9-540769-4277 (010)
Project Number: 8-0769
Project Director: Dr. Victor B. Cline, University of Utah, Salt Lake City, Utah
Period of research: 1 July 1969 through 31 December 1972

The Principal Investigator wishes to express deep appreciation to the Office of Education for its support of this basic research in the area of understanding and accurately judging others. In this Final Report we will attempt to summarize some of the major findings which have emerged from three-and-a-half years of research activity. The strategy, as proposed in the original prospectus, was to conduct a series of related studies using the technique of sound motion pictures of interviews and carefully developed judging instruments (as the basic measure of judging accuracy).

The Principal Investigator would like to acknowledge the invaluable contribution made by a dedicated group of research associates, assistants, and graduate students who contributed greatly to producing the many positive and significant outcomes of this research: Jon Atzet, Elaine Holmes, Robert Madsen, Dale Penprase, Joseph Pettit, Michael Shaw, Von Atkinson, and Carolyn Miller. In some cases some of the findings reported here resulted from Masters degree thesis projects.

CG 010 133

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Section I

Assessing the Validity of Verbal and Non-verbal Cues in Accurately Judging Others

Effective communication within groups as well as between pairs of individuals always requires a correct reading or interpretation of both the explicit verbal as well as the more covert non-verbal cues. The author and his associates have been engaged for some years in research (1964) in the area of person perception making use of sound motion pictures of interviews as the primary stimulus material and vehicle for their work. Brief motion pictures of interviews had one advantage of "freezing" or holding constant for experimental study a vast number of cues and information in a very dynamic context. The filmed interview always remained a constant no matter where presented or screened. It could be carefully studied and dissected by the experimenter at his leisure or seen and responded to by groups of judges anywhere, but it always remained the same, unchanging constant stimulus. This provided an ideal research setting to study the issue of cue validity, or attempt to determine what kinds of cues taken from a filmed interview contribute most to accurate judgment and appraisal of another.

A brief review of our basic methodology and research approach follows: Judges would typically be shown one or more films photographed in sound and color of an interview with a "standard other." These interviews were conducted by skilled interrogators who asked a fairly

standard series of questions (to insure equivalence over interviews). The interview, about 10 minutes in length, focused on personal values, personality strengths and weaknesses, personal interests, hobbies, recreation, aspects of life style, self-conception, impulse control, reaction to the interview, etc. The interviewer frequently probed beyond the initial question when responses suggested important data about the person that needed further inquiry or exploration. The pacing, order of question, techniques of probing, etc. had been developed following several years preliminary exploratory research. The aim was to get as much crucially vital data about the interviewee as possible in a 10-minute interview. It was found that those filmed interviews which went much beyond 10 minutes became increasingly difficult for judges to respond to with a high degree of attention. A kind of satiation factor set in even though there was some movement in the sense of frequent slow zooms in on the interviewee's face and out again to capture his total body and posture. Also, it was found that filmed interviews that were too lengthy, reduced the number of films that one could economically and reasonably present to a group of judges. Thus, with the 10-minute film it was possible to present a maximum of six in a two-hour period and have judges respond to these using specially developed judging tests.

After a filmed interview had been shown, the projector would be stopped and the subject-judges were required to fill out paper-pencil judging tests. These instruments were developed on the basis of intensive interviews, both with the person in the film and with five of his

close associates, friends, relatives, and others who have known him for at least five years. After the tests had been filled out the next filmed interview would be shown, after which the same type of judging instruments would be again administered to the subject judges.

The first judging instrument was referred to as the Behavioral Postdiction Test. Here the judge was required to guess or postdict how the person seen in the film (interviewee) behaved in everyday life. A series of 20 questions was developed with multiple-choice answers. For example:

12. When the interviewee (in the film) is in an argument, he usually:

___(a) becomes very sarcastic
___(b) leaves the room or area
___(c) uses profanity and obscene words
___(d) strikes or shoves his opponent

There was only one correct response. The other alternatives were very carefully chosen and constructed so as not to include even partially correct statements. A separate and individually tailored behavioral postdiction test was thus, necessarily, constructed for every film or interviewee. Other items tapped such areas as relations with women (men), social behavior, handling money, impulse control, etc. The experimenter of course knew how each interviewee characteristically behaved in real life because of the intensive interviews held with the subject and his family, close friends, business associates, etc., about his life style, personal history, etc. Only those behaviors which all agreed upon as being characteristic of him were used. Thus a separate schedule of items had to be built for every interviewee. The judging ability scores were

obtained by tallying the number of correct predictions/postdictions made for all of the filmed interviews used.

The Adjective Check List, a modification of Gough's ACL test (e.g., a list of 300 trait-adjectives) was developed. In this case the judges were required to postdict a filmed interviewee's self-ratings on the ACL. Twenty pairs of descriptive adjectives or traits were presented to the judge. Each judge was required to predict which trait (from every pair) the filmed interviewee and his five friends and associates consistently checked as being descriptive of the interviewee. The correct trait was the one which he and all or nearly all of the interviewee's friends and associates had chosen to use in describing or rating him. The incorrect trait was the one which had never (or almost never) been checked in describing or rating him. The correct and incorrect traits in each pair were matched as closely as possible for "social desirability" to prevent this response bias from influencing the judge in his decision. Example:

18. ___(a) Strong
___(b) Loyal

A number of other judging tests were developed which space prevents a detailed discussion and description of. They involved such things as tests of the judge's verbal and visual memory of the film's content, ability to predict the interviewee's personal values, etc., etc. Our various judging measures do have a moderate relationship with each other and our research (1961b) does suggest a modest amount of "general ability" in one's ability to judge others.

In studying the problem of what kind of cues or information presented

in the filmed interviews contributed most to accurate judgment and appraisal of others by our judges we discovered that some films (interviewee's) were relatively easy to judge and others very difficult. This varied on the continuum of "transparency" to "opaqueness." Thus, some individuals are very open, easily perceived, easily understood and accurately judged by nearly everyone--hence transparent. However, moving to the other end of the continuum we found some individuals or interviewees who projected very confusing cues. The more you studied and listened to them the more contradictory and confusing they appeared. It was as if they were playing roles, and one didn't know for sure who the real man was underneath. We referred to the more extreme cases of this type of individual as "opaque." We found that if an individual were either too transparent or too opaque we could not use the film. It would be akin to an item in a math test that was too easy (everybody passed it) or too difficult (everyone failed it)--it didn't discriminate so had to be eliminated. Two examples might illuminate this point. One day while filming our interviews in a downtown urban area we noticed a young man possibly 16-18 years of age, not too well-dressed, who in his general demeanor somehow suggested the image of a young "hood." Since he fitted into the age and socioeconomic category we were looking for, we approached him and offered him a small sum of money to be interviewed. He accepted and we immediately began the interview (with cameras running) wherein we learned that he was from an impoverished background. His father had died six years previously and he lived with his widowed mother and several younger siblings. From the interview we learned that he had

a warm and affectionate relationship with his mother, went to church every Sunday, loved his country, and was doing the best he could to support, at a very meager level, his family. We were all touched by his story. We gave him extra money for participating and bade him God-speed. Our later background study of him showed that he was a hardened, calloused criminal involved in many crimes of violence, attempted murder, rape, etc. However, he was a consummate role player and had the facility and ability to project an image very different from what he was like in real life. We could not use his film because almost no one could penetrate the false role of a "pleasant all-American boy down on his luck" which he presented in the film.

In another instance we made a film of a very attractive blonde young adult female who in the interview projected a great deal of "animal vitality." Nearly all judges, men and women, did very poorly in judging her as long as they were watching her presence on the screen. When the judges only heard or even read a transcript of the interview but did not see her, judging scores rose remarkably. When her image was on the screen the judges apparently paid little attention to what she said, or discounted it if they did hear. Her visual presence, powerfully and inaccurately cancelled out the content of her verbal message.

What all this suggests is that both verbal and non-verbal cues can be very deceptive and misleading. They may convey messages which obscure the truth, which trap and mislead, deliberately perhaps. The real issue may be how do we decode these false cues, (verbal or non-verbal) or

penetrate the mask or role of the opaque individual or more importantly tell whether the cues are false or true, and whether we are dealing with a transparent or opaque individual? The transparent person is easily deciphered and read by most who interact with him at both verbal and non-verbal levels. He's not our problem. One can imagine another type of person, neurotic and conflicted perhaps, who may project conflicting cues. But this could be another instance of transparency where the truth is that there is complexity and ambivalence. An example might be the child who both loves and hates his mother. Yet there is that other person who is truly opaque and may generate sufficient false cues to grossly trap and mislead us. Where considering marriage or hiring for a key company position is involved, this could be "fatal."

We might next proceed to examine some of the empirical research related to the cue validity problem to gain a greater understanding of what types of cues are most helpful or powerful in contributing to the accuracy of judgment and knowledge of one person by another. The first study to be cited is referred to as the "Cue Manipulation Research."

The purpose of this study was to systematically alter the input of information or cues available to judges about a particular "other" or subject and then determine what effect this might have upon judging accuracy scores. The standard procedure was to use the sound color films, which had already been developed with four judging tests. The first was the Behavioral Postdiction test already discussed. The second judging device was the ACL or Adjective Checklist test which was previously described. The third procedure, the Verbal Memory test, was a procedure

requiring the judge to show how well he remembered the content of the filmed interview he had just seen. A fourth was a Visual Memory test which involved having the judge indicate what had happened or what he had seen in the film. All experimental groups saw six filmed interviews and filled in the appropriate judging tests. Eleven experimental groups were organized consisting of 25 adults each, as follows:

Group 1 Saw and heard the standard set of six judging films and filled out all judging instruments.

Group 2 Were given only information about the age and sex of each interviewee in each film and then without seeing the film they filled out the judging instruments. This was a condition of minimum information.

Group 3 Read 12 revealing facts about each interviewee (in each film) such as occupation, education, sex, marital status, race, and so forth. Then on the basis of only this information, they filled out all judging instruments.

Group 4 Saw only the visual content of the film but heard nothing, they then filled out all judging instruments.

Group 5 Heard only the sound track, but saw nothing. They then filled out all judging instruments.

Group 6 Heard the sound track and saw 30 seconds of each interviewee's behavior and then filled out all judging instruments.

Group 7 Saw the visual content of the filmed interview and heard each interviewee's voice for 30 seconds only. They then filled out the judging instruments.

Group 8 Read a printed transcript of each interview. This was the

only information they were given. They then filled out all judging instruments.

Group 9 Read a personality sketch only about each interviewee. They then filled out all judging instruments.

Group 10 Read a personality sketch and also saw and heard films. They also then filled out all the judging instruments.

Group 11 Read a personality sketch of the interviewee, saw and heard films, and then filled out all judging instruments. Following this they additionally read the printed transcript of each interview and again filled out all the judging instruments.

In analyzing the results it was found, using an analysis of variance across all eleven groups and four judging instruments, F ratios were all significant at much better than .01 level.

The Newman-Keuls sequential range statistic was used to test the significances between individual group performances on each of the four judging tests across the experimental groups. For the Behavior Prediction test the groups clustered into three distinct levels of accuracy. Groups 2, 4, and 7 tended to be quite low in their accuracy. Groups 1, 3, 5, 6, and 8 were accurate to an intermediate level, and Groups 9, 10, and 11 were high. For the ACL test three very similar groupings also emerged; Groups 2, 4, and 7 being quite low and Groups 1, 3, 5, 6, and 8 being intermediate, and Groups 9, 10, and 11 being high. The differences between these clusters were significant. These results in summary indicate that what most interviewees say in response to interview

questions is far more important, as cues, than what they look like, what the voice sounds like, how they act or move (without the sound) all put together. Also, most importantly, having an accurate stereotype like knowing 12 relevant facts about the person such as race, religion, political inclination, marital status, etc. gives a judge (in one instance) as accurate a judging score as hearing and seeing the entire filmed interview. This is true for at least one judging device and that is the Behavioral Postdiction test. The most accurate groups of all were those groups who read the personality sketch and/or read the printed transcript of the interview. However, it should be noted that the sound quality of some of the films was only fair, which would permit anyone reading a written transcript of the interview to have an advantage in probably having more information available to him.

With regards to the Verbal Memory judging test, Groups 1, 5, 6, 8, 10, and 11 achieved a high level of accuracy, while Groups 2, 3, 4, 7, and 9 obtained low scores. This finding is easily explained on the basis of exposure of the subjects to the verbal content of the interview (orally or reading a verbatim transcript). Also, those who read the interview transcript achieved a significantly higher score than those who just heard it.

With regards to the Visual Memory test, Groups 1, 4, 6, 7, 10, and 11 achieved relatively high scores, having been exposed to the visual content of the films. Even a 30-second exposure proved of major value on scores achieved on this instrument. However, a longer visual exposure such as was obtained in Groups 10 and 11 substantially raised the total

Table 1

Analysis of Variance of Judging Accuracy Scores Across Eleven Groups

(Analysis Broken Down by Judging Instrument)

	Source	DF	MS	F	P
Behavioral Postdiction	Group	10	1842.86	56.07	.01
	Ss/Group	264	32.87		
	Total	274			
ACL Test	Group	10	1168.56	84.74	.01
	Ss/Group	264	13.97		
	Total	274			
Verbal Memory	Group	10	14071.766	119.76	.01
	Ss/Group	264	117.49		
	Total	274			
Visual Memory	Group	10	7068.78	46.86	.01
	Ss/Group	264	150.85		
	Total	274			

Table 2

Comparison of Treatment Totals via Individual Degrees of Freedom Test[a]

(Behavioral Postdiction Test)

Order of Absolute Totals by Group										
4	2	7	8	3	1	6	5	10	11	9
984	986	990	1141	1199	1226	1231	1249	1478	1545	1567

Difference Matrix

	4	2	7	8	3	1	6	5	10	11	9
4	--	2	6	157**	215**	242**	247**	261**	494**	561**	584**
2		--	4	155**	213**	240**	245**	259**	492**	559**	581**
7			--	151**	209**	236**	241**	255**	488**	555**	577*
8				--	58	85	90	104	337**	404**	426**
3					--	27	32	46	279**	346**	368**
1						--	5	19	252**	319**	341**
6							--	14	247**	31 **	336**
5								--	233**	300**	322**
10									--	67	89
11										--	22
9											--

[a]Significance calculated via the Newman-Keuls Sequential Range Statistic.

MSErr. = 32.87 * = signif. at .05 level

df = 264 ** = signif. at .01 level

K = 25 N = 25

accuracy score still more on this measure. It was found that the Behavior Postdiction and ACL tests were intercorrelated .30 while the two memory tests were correlated .25. Neither the postdiction nor trait judging tests were significantly related to either of the memory tests. While this study is certainly limited in the degree to which we can generalize from it, it certainly does emphasize the importance of verbal cognitive material and tends to play down such things as style, appearance and the many so-called subtle visual cues which one might guess would be important when making judgments about another person. One thing noticed in conducting these series of researches over the years has been that many times seeing the person you are interviewing can actually interfere with or reduce the accuracy of one's judgment. Thus while it may be possible for a certain judge to be particularly sensitive to style, movement and expressive behavior and this kind of thing, and these may contribute significantly to his judging ability, still it is possible that many interviewees are actors or role players who tend to project false images of themselves to others which could potentially impair one's judging accuracy. A further implication of this study is that descriptive demographic or what one might call stereotypic information about the person such as his religious preference, racial background, social class status, ethnic derivation, occupation, marital status, and so forth, are of great importance in understanding and making predictions about a person. While the word "stereotype" is almost regarded as a "dirty word" when judging people, especially to social psychologists, our data here suggest that it is most important and if used in conjunction

with idiosyncratic data, serves to yield powerful and important trends or data about a person, upon or about which quite potent predictions might be made about his future behavior as well as present functioning. Further work obviously needs to be done in this area. But it certainly points the way to further fruitful kinds of research endeavors.

The next study is referred to as the "feedback in judging accuracy training research."

Two hundred and seven subjects were divided into eight groups of twenty-three which formed all possible combinations of three dichotomous variables: (a) feedback[1] - no feedback, (b) summary[2] - no summary, (c) films[3] - no films; plus a ninth group which received no feedback, no summary, saw no films, but was told the age and sex of the stimulus person. A description of the treatments imposed upon the groups is as follows:

Group 1 Films, summary, immediate feedback

Group 2 Films, summary, no feedback

Group 3 Films, no summary, immediate feedback

Group 4 Films, no summary, no feedback

Group 5 No films, summary, immediate feedback

---

[1]Here a judge was given immediate feedback about his judging accuracy following each response to each item.

[2]Here a judge was given a brief written summary and demographic information about the person he was judging.

[3]Here a judge saw and heard the complete filmed interview, which he later made judgments about on four tests.

Group 6 No films, summary, no immediate feedback

Group 7 No films, no summary, immediate feedback

Group 8 No films, no summary, no feedback

Group 9 No films, no summary, no feedback, but were told the age and sex of the stimulus person

The films were shown and questionnaires administered to each of the above groups in the prescribed manner.

Films - no films: The results of an analysis of variance showed that the no-film groups did as well as the film groups on only one judging instrument, the Postdiction Test. Here the mean accuracy scores of the groups were not significantly different from each other ($F = 3.70$; $df = 1, 176$; $p \geq .05$). Significant differences in mean accuracy scores were found for the remaining instruments:

ACL: $F = 4.77$; $df = 1, 176$; $p \geq .05$

Verbal Memory: $F = 1678.64$; $df = 1, 176$; $p \geq .001$

Visual Memory: $F = 29.47$; $df = 1, 176$; $p \geq .01$

Summary - no summary: The analysis of variance showed that summary groups were superior to no-summary groups on all questionnaires:

PD: $F = 11.29$; $df = 1, 176$; $p \geq .05$

ACL: $F = 51.69$; $df = 1, 176$; $p \geq .05$

Verbal Memory: $F = 7.39$; $df = 1, 176$; $p \geq .01$

Visual Memory: $F = 21.16$; $df = 1, 176$; $p \geq .001$

Feedback - no feedback: The analysis of variance showed that feedback groups were not significantly more accurate on the Verbal

Memory Questionnaire (F = 3.54; df = 1, 176; $p \geq .05$). Significant differences were found for the remaining instruments:

PD: F = 19.11; df = 1, 176; $p \geq .01$

ACL: F = 18.62; df = 1, 176; $p \geq .01$

Visual Memory: F = 6.51; df = 1, 176; $p \geq .05$

Significant Interaction: Two-Factor

Films-feedback: This interaction term was not significant for the Verbal Memory Questionnaire (F = .694; df = 1, 176; $p \geq .05$). The interaction was significant for all other instruments.

ACL: F = 4.19; df = 1, 176; $p \geq .05$

Visual Memory: F = 14.86; df = 1, 176; $p \geq .01$

Summary-films: This interaction term was significant on the Post-diction test only (F = 19.11; df = 1, 176; $p \geq .01$).

Significant Interactions: Three-Factor

The three factor interaction term was significant for all instruments:

PD: F = 11.62; df = 1, 176; $p \geq .01$

ACL: F = 14.02; df = 1, 176; $p \geq .01$

Verbal Memory: F = 7.51; df = 1, 176; $p \geq .01$

Visual Memory: F = 18.68; df = 1, 176; $p \geq .01$

The existence of significant interaction terms means that care must be taken in the interpretation of any one factor. In order to assess the effect of a factor, the level of the other factor(s) must also be known. The following hierarchies of accuracy (see Table I) can be arranged, with the conditions most conducive to accurate scores at

the top, and those least conducive on the bottom. These factors can be assessed for conditions in which only one factor exists at a time and when two or more are in combination.

Providing some source of information does improve accuracy on the ACL and Visual Memory Questionnaires, but the importance of the kind of information provided is not the same for the two instruments (on ACL, feedback is the most powerful discriminant; on Visual Memory, films are the most powerful discriminant). On the other hand, accuracy scores would not be improved on the Postdiction and Verbal Memory Questionnaires merely by adding new information. In particular, scores on the Postdiction test would not be improved by providing the information found in the films, and Verbal Memory scores would not be improved by providing feedback of correct answers. Therefore, whether or not additional information will increase accuracy scores depends on the kind of information provided and the judging instrument that is used as the dependent criterion.

Some of the implications of this research are that, as in the case of the Behavioral Postdiction test, it is not necessary to see the person to make the most accurate judgment about him. Getting information about him in the personality sketch, plus the feedback seem to be quite potent in producing high accuracy scores. In the case of the ACL test the films contributed some to accuracy, but feedback and the personality summary with its demographic (or "stereotype" type) data contributed more. In the case of the Verbal Memory test the films were more important here, for the very obvious reason that it is only

by seeing and hearing the film that one gets a complete knowledge of what the person says in response to the questions. And since the judging test is taken directly from a transcript of the film, this is the only condition in which the judge is able to get specific data about what was said, so these results are to be expected. In the case of the Visual Memory test the films as might be reasonably expected are also the most important contributor of information and data for the same reason. This study also gives somewhat indirect evidence that people can be taught to improve their judging scores by providing immediate feedback to a judge on whether he is correct or not in his individual judgment about a person. This is a very powerful kind of technique in helping our judge to obtain high accuracy scores.

Thus, in summary, accurately perceiving and judging another is a complex task which is affected by many variables. Cues can be covert or overt, verbal or non-verbal--but in either case ease in decoding them can range greatly on a continuum of difficulty depending on the transparency or opaqueness of the person being judged. Additionally, people can be trained to be better judges of others by such techniques as giving the judge feedback about his judgments. And at least with the research methodologies used by the present investigators it was found that the verbal content of the interview (heard or read) plus knowledge of some demographic background data about the subject contributed more to judging accuracy than any other factor including "seeing" the subject.

Table 3

Intercorrelations of Scores on Four Judging Instruments

All Possible Intercorrelational Combinations by Groups						
Group	PD[a] x ACL	PD x AM	PD x VM	ACL x AM	ACL x VM	AM x VM
1	.42	-.24	-.04	.10	.00	.17
2	.39	-.04	.03	.05	.05	.48
3	.37	-.18	.01	.02	.13	.45
4	.20	.04	.09	.08	.02	.22
5	.11	-.09	-.08	.02	-.10	.18
6	.37	.01	.00	.06	.12	.31
7	.30	.04	.04	.07	.02	.11
8	.25	.01	.01	.10	.05	.15
9	.51	.04	.16	-.13	.04	.26
10	.31	.25	.11	.17	.16	.18
11	.09	.05	.22	.10	-.03	.22

[a]PD = Behavioral Postdiction
ACL = Adjective Checklist Test
AM = Verbal Memory
VM = Visual Memory

Table 4

Hierarchies of Conditions Conducive to Accuracy

	Individual Factors	Combination of Factors
PD	1. Feedback	1. Feedback + summary
	2. Summary	2. Feedback + no summary
		3. Summary + no feedback
ACL	1. Feedback	1. Feedback + summary - films
	2. Summary	2. Feedback - summary - no films
	3. Films	3. Feedback - no summary - films
		4. Feedback - no summary - no films
		5. No feedback - summary - films
		6. No feedback - summary - no films
		7. No feedback - no summary - films
		8. No feedback - no summary - no films
Verbal Memory	1. Films	1. Films + summary
	2. Summary	2. Films + no summary
		3. No films + summary
		4. No films + no summary
Visual Memory	1. Films	1. Films - summary - feedback
	2. Summary	2. Films - summary - no feedback
	3. Feedback	3. Films - no summary - feedback
		4. Films - no summary - no feedback
		5. No films - summary - feedback
		6. No films - summary - no feedback
		7. No films - no summary - feedback
		8. No films - no summary - no feedback

Section II

The Effects of Money, Anticipation, and Competition Upon Judging Scores and Self-Reports of Motivation Arousal

Procedure

Selection of Sample

149 adult male and female S's were drawn over a period of eight months from the introductory psychology, abnormal psychology and introductory sociology classes at the University of Utah and Weber State College. Most students in the aforementioned classes had an opportunity to participate. In some cases, with instructor permission, the entire class was used. In other cases professors encouraged their students to participate and gave extra credit to those who did so during evening sessions.

Without their knowledge Ss were preassigned one of seven experimental conditions which were six treatments and a control. Each of the six treatment groups was assigned one of three variables. These were money, anticipation of results and competition between sexes. Insert #1 summarizes the experimental design as it tells the approximate date and place of completion for each group involved.

Procedures Followed for Each Group

By means of a standard information sheet handed to all Ss at the beginning of the experimental session, the E read the procedures to be

Insert 1

Experimental Design*

Experimental Groups	Film 01	Film 09	Completion Date & Place
MONEY---30¢ given for each correct answer given on the behavior postdiction test. The money was shown before the film began.	XX		May Utah
		XX	May Utah
ANTICIPATION---Ss were informed that they would receive their test scores and papers after the film along with a group mean.	XX		June Utah
		XX	June Utah
COMPETITION---Ss were informed that the sexes differ in their ability to judge. They were to compete; a mean for each sex was given.	XX		August Weber
		XX	August Weber
CONTROL---Ss were only shown the films. They were not given any treatment condition along with the films.			July Utah

*An X within a cell indicates that for the film the particular group in question received the treatment condition explained.

A blank cell indicates that for the film the particular group in question received only the film, not the treatment condition.

Film 01 was always shown first while Film 09 followed. After having seen each film Ss were requested to fill out a 21 question behavior postdiction test which was attached to the instruction sheet.

Attached to each instruction sheet and behavior postdiction test paper was a questionnaire. Ss were also required to complete the questions on that paper after having viewed a film and taken its test. This questionnaire contained, along with irrelevant data, three Likert-type questions. These were designed to assess by subjective report how hard the Ss tried, concentrated and thought during a particular film and test session. Insert #3 gives the three questions that were used.

Insert 3

Motivation Questions on Questionnaire

1. How hard did you try to get all the answers correct?

X	X	X	X	X	X	X
Not at all	Not Much	Not too much		Tried	Tried quite a bit	Really tried

4. How much did you concentrate during the session?

X	X	X	X	X	X	X
A great deal	A lot	A little		Not too much	Not much	Not at all

7. How deeply did you think through the questions?

X	X	X	X	X	X	X
A great deal	A lot	A little		Not too much	Not much	Not at all

Having completed the foregoing for Film 01, Ss repeated the same procedure for Film 09. That is, they viewed the second film, completed the test for that film and questionnaire they received before the

film began, and awaited final instructions by the E. In this case, after the second film, Ss were told they could leave.

For the control group the above procedure comprised the sum total of each Ss experience. Treatment groups, however, received an added variable depending upon the group involved. Ss in the 01 money group, for example, were told through an added instruction sheet, inserted between the instructions and the test plus questionnaire and read by the E, that for the 01 film they would receive 30¢ per correct response on the postdiction test. The promised money in dollar bills, quarters, dimes, and nickels was laid upon the E's desk at the rear of the experimental room after having been shown the Ss. Insert #4 is a facsimile of the added instructions given the money groups.

Insert 4

Money Group Instruction Sheet

On this particular test you will receive 30¢ for each correct answer given. Since there are 21 possible responses, you may receive up to $6.30 for your half hour's work along with the experimental point for the session. It might be added that provision was made in the original grant for the allocation of this money. You, therefore, need not feel guilty or inhibited in any way in accepting this stipend.

As soon as you finish, you may bring your test to me and immediately upon its scoring you will receive the appropriate amount. Questions?

As this added instruction sheet explains, Ss obtained an immediate award. After each "judge" completed his test and questionnaire, he took his papers to the E who scored the test and gave the S the proper amount of money for his score. Ss did not see their test papers after their correction. As soon as all Ss completed their

tests and received their awards, they viewed the second film (09), filled out its test, completed the attached questionnaire and left the room.

The 09 money group followed a reverse procedure to that followed by the 01 money group. These Ss received the experimental instructions, saw the 01 film and filled out afterwards the 01 test and questionnaire attached to their instructions. They were then given the special instruction sheet about the possibility of 30¢ per correct answer along with the 09 test and accompanying questionnaire. Ss in this group, therefore, were given money for their responses on the 09 test instead of the 01 test.

The 01 anticipation group followed the procedures explained above for the 01 money group with one exception. Instead of 30¢ per correct answer, they, the 01 anticipation Ss, were promised the results of their individual tests along with a group mean for the test in question. Insert #5 is a facsimile of the instructions given the anticipation groups.

Insert 5

Anticipation Group Instruction Sheet

On this particular test you will receive the results. By comparison of your responses with those which are correct you will see where you accurately or inaccurately described the individual in the film. A mean score will also be compiled for the group as a whole and you may see how far above or below this mean your score falls.

As soon as you finish, you may bring your test to me and immediately upon its scoring it will be returned for your consideration. We ask only that you return it before you leave. Questions?

These Ss did receive their test papers back after correction. They would, after seeing the 01 film, fill out their test and questionnaire, turn these in to the E at the rear of the room and receive the scored sheets back in a few minutes with the corrected answers of the questions missed circled in red. This was done only after all Ss had finished. That is, corrected test papers were taken by the E and returned to the Ss after everyone had completed.

The 09 anticipation group followed the same basic format used for the 01 anticipation group. They received, however, their results and the group mean after the second film. In this respect they were identical to the 09 money group in that their treatment, anticipation of results, came for the 09 film instead of for the 01 film.

The competition groups were very similar to the other treatment groups above. The major difference between these two groups and the other four treatment groups was that the payoff for Ss within them was a sex comparison rather than money for correct responses or anticipation of results. This sex comparison was a male and female mean for the test given under treatment conditions. Other than this difference in the variable being tested, the 01 and 09 competition groups were basically identical to the 01 and 09 groups for the other two treatments. The only other procedural difference for these competition groups was that their instructions were given orally by the E while the Ss did not have an instruction sheet about their treatment variable with which they could follow along as the E read to them. Insert #6 is a copy of the explanations given these Ss.

Insert 6

Competition Group Instructions

It has generally been shown that the sexes differ with regard to their ability to judge another. Usually males judge one of the sexes better than they judge the other, and females generally judge one of the sexes better than the other also. We would like you to compete with each other, therefore, to see which of you---the boys or the girls---are better judges of the person in this particular film.

Do not consult with each other. Just try as best you can to do well. After everyone has finished and the tests are corrected a mean for the two sexes will be given so that you can tell who did the best. Are there any questions?

Tests and Computations

Since after each film in all groups the Ss took the behavior postdiction test and completed a questionnaire that contained three "motivation questions," the E had four scores from all participants. These were, once again, the score each S obtained from taking the test and the three scores each S received from marking his response to the "motivation questions" where possible scores ranged from -3 to +3 on the seven point scale. In each of the seven experimental groups all four scores were treated separately as they were totaled across Ss and listed under their separate headings. These headings were named after the principal word of each motivation question (try, concentrate, and think) while the word scores served for the postdiction test scores.

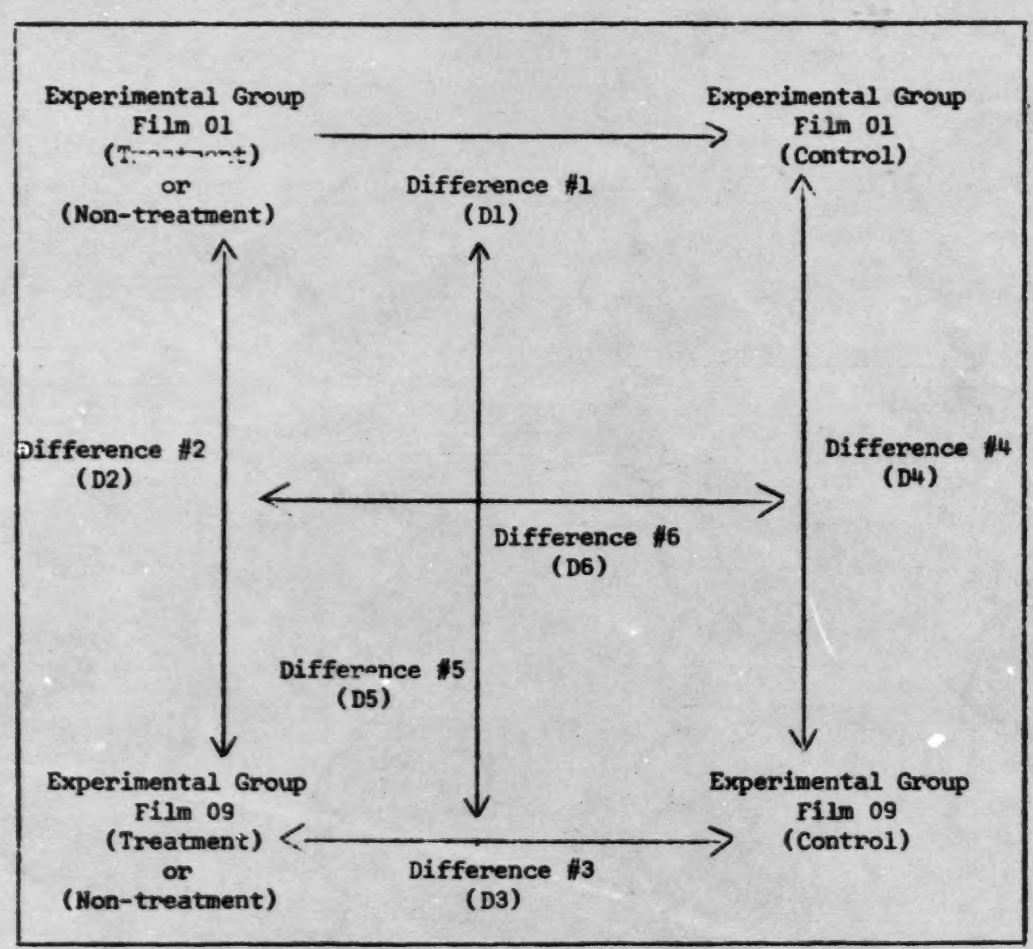
A fifth measure was also computed by combining the data from the three "motivation questions" into one "combined questions" sum. This was in essence the average of the three separate questions, and it was labeled under the title, combined questions. Each group had,

therefore, five measures for which raw score totals, means and sums of squares were computed to be used for inferential comparison.

Several areas of comparison between the control and experimental groups were hypothesized. These differences are shown in Insert #7.

Insert 7

Possible Differences Between Means



It is easy to see that there were various differences that could have been tested for significance. It should be noted, however, that only three of these differences were seen to be important to this particular experiment. These relevant differences were: (1) The difference between the control and an experimental group that received a treatment condition for the film used for comparison. (D1) or (D3) would be the two differences spoken about here. (2) The difference of change within a treatment group compared with this same change within the control group. This would be a comparison of (D2) and (D4). (3) The difference of change from film to film within experimental treatment groups. This would be an analysis of (D2).

The significance of difference #1 above (D1 or D3) was tested by a Fisher's t for uncorrelated means. The sum of the raw scores and the sum of the squares of these scores lead easily, along with the means concerned, to a comparison between the experimental and the control groups on the five measures. To test difference #2 above was much more difficult. This test first of all required the individual subtraction of the five means obtained from the 09 films from their corresponding means obtained from the 01 films. The resultant loss or gain for each measure in the treatment groups was then compared with a similar loss or gain which occurred from film to film within the control group on the same five measures. A Fisher's t was then employed again here to test the significance of the comparison. Finally, a Fisher's r was determined for scores from the 01 film measures and scores from the 09

film measures. With such a coefficient of correlation several ideas could be formed and used to discuss various implications surrounding difference #3 above or (D2) in the insert.[1]

Results and Discussion

Graph #1 summarizes the computations made testing the significance of D1 and D3.

Graph 1

Results of Testing for Significant Differences Between a Control and Three Treatment Groups, Each with Two Films

Treatment	Film	Try Question	Concentrate Question	Think Question	Combined Questions	Scores
MONEY---30¢ per correct answer given.	0_1					Yes .05
	0_9	Yes .01	Yes .05	Yes .01	Yes .05	
ANTICIPATION--- of correct answers and relative ranking.	0_1					
	0_9					
COMPETITION--- between males and females on scores.	0_1	Yes .01				Yes .05
	0_9					

*Yes in one of the cells indicates that a significant difference between the treatment mean and the control mean did exist at the level indicated by the number following it.

A blank cell indicates an insignificant difference between the treatment mean and the control mean for the measure in question.

[1] Hereafter the subscripts (D1), (D2), etc. will be used in reference to the differences between means either in comparisons between treatment groups and the control or within the treatments from film to film. The reader is counseled to refer frequently to Insert #7 if it becomes confusing to follow the text in the next section on results.

A quick glance over this chart shows that no experimental condition was consistently significant with either film across the four questionnaire measures and the scores from the postdiction test as well. Only the 01 competition group was significantly different from the control group with respect to the test scores and at least one of the measures from the questionnaire. This treatment group was significantly different on the "try question" and the test scores at the .05 level of significance.

Perhaps this is more important that if all of the motivation questions had been significantly different for this particular group. It is possible that only the "try question" was a valid instrument due to the fact that it was the first question to be answered. The other two motivation questions on concentration and thinking possibly became corrupted by the Ss response to the other items. If the Ss did have a "response set" to the questionnaire items in general then these two questions and the combination of the three would have suffered. There is no way to tell from this study if the Ss did in fact behave with such a particular "set." Unfortunately the 09 competition Ss failed to parallel the 01 competition Ss. Not even the "try question" let alone any of the other four measures were significant for this group.

The E believes that another variable or combination of variables were operating within this 09 competition group to obscure a possible result that would have been significant. Such belief is based solely on observations of the group used. It consisted of an August afternoon class of Weber State students. The day was hot and the proximity to being out of school might have created an apparent apathy that seemed to

exist toward the E and the experiment. The Ss seemed uninterested and "put out" to compete with one another. The 01 competition group, on the other hand, was also an August class of Weber State students. This group was run, however, in the morning hours well before lunch. The Ss seemed much more "alive" to the E and very interested not only in the experiment but also in the competition requested. Such an observation is only subjective on the part of the E, however, for no data was collected to ascertain the truth of this speculation. Perhaps the fact that the 01 treatment group, which showed significance on two measures, was compared with a control taken at the University of Utah but also from a morning class might lend credence to the above assumption that time of day could have been an important factor. A morning class compared with a morning class produced significance while an afternoon class compared with a morning class did not. Is the time of day such an indispensable factor? Further research is needed to determine this.

The two money treatment groups showed even more interesting results than those run on competition. While the 01 group used in the evening hours of May at the University of Utah was not significantly different at the .05 level on any of the measures, it was, interestingly enough, significantly different from the control (run in the morning hours during a class period) on the scores Ss obtained. The 09 group, for a contrast, was consistently different at or beyond the .05 level on all of the three questions as well as their synthesis. Surprisingly, however, this group failed to reach significance with respect to the test scores involved.

In this case it is difficult to question about a "response set" that invalidated the concentration and think questions. A very real significant difference between the control and the 09 money group did exist with respect the questions on the questionnaire. The Ss reported themselves as trying, concentrating and thinking more than the control Ss reported themselves. This group's failure to score higher, though, indicates that these 09 money Ss were either more liberal in their self-appraisal or that their heightened "motivational state" did not function to help them raise their test scores. Both of these hypotheses could be correct but the E feels that the former is less plausible than the latter.

It is possible, of course, that one or more other variables might have obscured the experimental effect expected. As questioned with the competition groups, for example, could it be feasible that it is easier to become motivated at one time of day as opposed to another? It is possible that the control group, arbitrarily taken from a class, was not representative of college Ss in general and unlike the other Ss used because they were enlisted from a pool of students that agreed to participate for extra credit in their respective courses? If so, could such a difference in our subject pool have accounted for the 09 money Ss that reported themselves as being highly motivated but might not have been as measured by their score achieved? The possibilities to explain the diversity of results within these two money groups seem endless. Further and more controlled experimentation appears essential to resolve some of them.

Both the money and the competition treatments with their gross inconsistencies are contrasted with the two anticipation groups which showed no significant difference on any measure when compared with the control. At least the consistency across these groups is refreshing. It was thought, however, that the anticipation of results would produce a marked change in all of the measured used for comparison. Such expectancy seemed eminently true if but a few of the Ss in each treatment group competed with themselves and/or others so that they might say to each other or within that they had scored so much higher than the group as a whole or the person that was sitting next to them. This occurrence is typical with the E's acquaintances, many of whom think deeply and try very hard on tests so that they can have the highest score or one that is better than that of another. If the Ss did in fact do this, no measurable results contributed to significance at the .05 level.

Once again it might be noted that both of the anticipation groups were run in evening sessions that volunteered to come at the end of a day to receive extra credit. They were being compared, as the other treatment groups, with a control group that was taken during class time in a morning class period that the instructor, not the students, had offered for the experiment. If the time of day and/or original motive of participation in the experiment were relevant variables, as has been hypothesized, then certainly these factors could be important with the two anticipation groups as well as with those mentioned earlier.

There remains, of course, the real possibility that there really

is no difference either for subjective reports of objective scores in anticipating one's results from a judging test under the conditions used. By the E's observations, however, these two groups seemed more motivated than the 01 competition group which did reach significance on two measures. These two groups did not, however, seem more aroused than either money group which also reached significance on several measures. Since not one statistical difference was found for a single measure in either of these treatment groups, the E, once again influenced by what he saw, thinks that possibly the motivation was present for at least many of the Ss but that they neither reported such nor scored higher on the post-diction test to help prove it. Could the anticipation for results have created differing "states of anxiety"that hindered Ss in their performance? Could the time of day or motive for participation have masked a real difference that might have been present? Once again it seems imperative that further and more controlled experimentation is necessary to investigate these questions.

The computations and logic behind the use of D6 were much more difficult than the tests and reasons behind the use of D1 and D3 as measures for inferential comparison. As stated in the procedure section, D6 was a test of the net change within the experimental treatment groups compared with a similar net change which occurred within the control group. It was assumed that all 01 treatment groups would have higher 01 film means on all measures than the 09 film means which followed them but did not receive treatment for the second film. It also seemed logical to presume that all 09 treatment groups which had received treatment on

the 09 film instead of the 01 film would have higher means on the second film. Following this assumption the second film measures were subtracted from their corresponding first film measures. The results from this procedure should have shown losses for all measures from 01 treatment groups and gains for all measures from 09 treatment groups. In other words, 01 treatment groups would have a minus sign before the difference for each measure between films and 09 treatment groups would have a plus sign before the difference for each measure between films.

This was not the case. Graph #2 shows that only about fifty percent of the difference scores between films (D2 in Insert #7) were actually in the direction expected. These results indicated that the treatments were not being measured as having a consistent effect in motivating the Ss. To see if the magnitude of any chances were significant, D6 was tested.

Results showed that only one of the D2 scores significantly deviated from the corresponding D4 score which occurred in the control. This one significant measure was the "try question" for the 09 money group, and it can be seen from the above graph that the D2 difference for this measure was the largest of all D2's involved. All other comparisons of D2 with D4 proved insignificant at the .05 level.

This finding might seem surprising when one sees in the graph that other D2's were almost as large as the +.85 that led to significance. It would also seem strange if it were confused with findings mentioned earlier. This D6 test of significance should not be confused, however, with what has been said earlier about those statistical results reported

Graph 2

Differences Between Means Within Experimental Groups*

Treatment Group	Film	Try Question	Concentrate Question	Think Question	Combined Questions	Scores
01 Money	01	2.00	1.90	1.75	1.88	10.35
		-.40	-.10	+.10	-.13	+.10
	09	1.60	1.80	1.85	1.75	10.45
09 Money	01	1.57	1.58	1.42	1.50	9.89
		+.85	+.79	+.68	+.78	+1.31
	09	2.42	2.31	2.10	2.28	11.26
01 Anti-cipation	01	2.11	1.89	1.78	1.93	10.11
		-.45	-.06	-.06	-.19	+1.04
	09	1.66	1.83	1.72	1.73	11.15
01 Compe-tition	01	1.65	2.05	1.80	1.83	9.25
		+.20	-.30	+.10	.00	+1.20
	09	1.85	1.75	1.90	1.83	10.45
09 Compe-tition	01	2.42	1.63	1.68	1.91	10.42
		-.69	-.11	-.11	-.30	-.10
	09	1.73	1.52	1.57	1.61	10.32
Control	01	1.40	1.50	1.70	1.53	8.65
		-.10	+.10	-.28	-.08	+1.35
	09	1.30	1.60	1.45	1.45	10.00

*The first figure in each cell is the first film mean for the treatment group or the control on the particular measure in question.

The second figure in each cell with a plus or minus sign is the difference between the first and second film means listed when the 09 film mean is subtracted from the 01 film mean.

The third figure in each cell (on a vertical line below the first) is the second film mean for the treatment group or the control on the particular measure in question.

in Graph #1. That diagram gave the significance of difference for direct comparisons between experimental groups and the control for each separate treatment film. Looking at the chart of differences (Insert #7), the reader will see that those earlier results and discussion were concerned with D1 for the 01 treatment groups and D3 for the 09 treatment groups. D6, on the other hand, is entirely different in nature from these other differences.

D6 is a comparison across measures of the net change from film 01 to film 09 in an experimental group with the corresponding net change across measures from film 01 to film 09 in the control group. This comparison was made in lieu of testing the significance of difference from film to film within treatment groups (D2 alone), because it was possible that the films were differently motivating in and of themselves. If this were the case, then Ss would conceivably get higher scores on any or all of the measures just because one of the persons interviewed in the films was more agreeable, similar, or arousing to the Ss than the other person interviewed. This possible difference between persons judged would not bother the D1 or the D3 tests for significance because the hypothesized "likeability" factor would be a constant that helped or hindered all Ss in the control and treatment groups alike on that specific film. D1 would still, therefore, be a test of the motivation treatment alone if 01 were the more "arousing" person interviewed, and D3 would be the same if 09 were the more exciting person filmed.

If the 01 filmed subject was more motivating, however, D2 would be a test of a multiple variable, the latent difference in films and the treatment imposed.[1] For this reason D6 was used in place of D2 as a test for significance of difference between means from one film to the next. It took into account the possible difference one filmed person would "excite" the Ss more than the other.

Recalling that only one D6 was found to be significant, it seems imperative that another conclusion be drawn. For we must suspect that even though some of the other direct comparisons of the difference between treatments and the control were strong and significant, the net change within Ss from film 01 to film 09 was very small and almost invariably insignificant. If this was a failure of the treatments in arousing the Ss or if it was a failure due to other variables that might have masked a real and present effect cannot be ascertained from the data. With D6 as with D1 and/or D3 we can only say that most of the data indicated a negligible difference between groups. The reasons for such an occurrence are to be found by more research.

A final test proved interesting to one of the hypotheses given earlier, however. The hypothesis in question, expressed several times already, states that perhaps induced motivation arousal creates "states of excitement" that differ within Ss as they act to help some while they

---

[1] Graph #2 shows that when compared with the control, ten of the fifteen measures in the three 09 treatment groups that had nothing in the way of treatment for the first film were higher than similar 01 film measures in the control. This difference (D1 in Insert #7), although not significant, was an indication that film 01 might have been a little more "motivating" to the Ss than the 09 film.

hinder others. The computations which lend some validity to this assumption were correlation coefficients computed between the measures on film 01 and the measures on film 09. Such coefficients are between the same means used earlier to obtain the D2 difference within groups. Graph #2, used twice before, also shows these means.

A positive correlation for these two scores would indicate that those who scored high on the first test and questionnaire measures also scored high on the second test and questionnaire measures. It would show besides that those who scored low on the first film measures also scored low on those of the second film. This coefficient would not signify that either high or low scorers on the first film did better on the second film, let alone significantly better. Nor would such a statistic indicate that a treatment condition caused the scores of the second film to be either higher or the same as they were for the first film. A positive correlation would indicate a trend, however, that showed that Ss were basically in the same general ranking from film to film.

A negative correlation, on the other hand, would indicate that those who scored low on the first film measures scored high on the measures of the second film. The converse, of course, would also be true. A zero or near zero correlation, to complete all possibilities, would indicate that Ss were randomly different from film to film. That is, while some Ss were high and others low on the first film measures, these same Ss were likely to be anywhere in the ranking of the second film results.

Graph #3 shows the correlations between films in the treatment

Graph 3

Correlations Between Treatment and Non-Treatment Measures Within Groups*

Groups	Treatment	Try Question		Concentrate Question		Think Question		Combined Questions		Scores	
Non-Treatment	Film	01	09	01	09	01	09	01	09	01	09
Money	01		-.36		-.28		-.18		-.26		.00
Money	09	-.24		+.43		-.04		-.02		+.47	
Anticipation	01		-.02		+.13		-.10		-.02		+.02
Anticipation	09	-.09		-.38		-.26		-.24		+.02	
Competition	01		-.06		+.23		-.31		+.02		+.09
Competition	09	+.11		-.27		+.24		.00		-.22	

*Each figure represents, for the measure in question, the coefficient of correlation between the film that received the treatment condition and the film that did not

groups on the five measures. Eighteen of the thirty correlations are negative and only one of those positive is significant. Perhaps more important seems the fact that most of the coefficients are low and not more positive. If they were a little more positive, then a consistent trend, as mentioned above, would seem to show that the introduction of a treatment condition at least kep the Ss ranked relatively the same

from film to film. Such would apparently say that the treatments uniformly raised or lowered Ss scores from one test to the next depending upon which film received the treatment condition. The fact that the correlations are small and more negative than positive indicates, if but slightly, that the three treatments either had little or no effect to speak of, or that the effect that they produced was mixed and led to contradictory results. If the latter is the case, further experimentation is called for once again to find those conditions that motivate and the magnitude at which such conditions must be present to produce a uniform effect upon most Ss.

Summary of Results

This experiment was designed to test the hypothesis that various treatments might produce significant effects on Person Perception test scores and subjective reports of motivation arousal as measured by three questions on a questionnaire. Money, anticipation of results, and competition between sexes were used as the treatment conditions under observation. The following is a summary of the findings of three tests for significance of differences:

(A) When compared with a control, money and competition treatments produced significant differences at or beyond the .05 level on:
(1) two out of four test means; and
(2) five out of twenty questionnaire derived means.

(B) When compared with the net change of scores from film to film within a control, only the net change of score from film to film for the "try question" of the 09 money group proved significant.

(C) Correlations between first and second film means within groups were small and generally negative.

Two hypotheses were given to explain this data: (1) It was hypothesized that the disparity between groups regarding the original motive Ss had in participating as well as the different time of day groups were run could have invalidated the study and any possible effects, present but masked from the treatment conditions. (2) It was thought that possibly even with more controlled experimentation Ss might still differ widely in their response to the treatment procedures. That is, some might score higher on the test or questionnaire measures as others scored lower on either or both of the same.

Section III

Social Stereotyping and Its Relationship to Accuracy in Person Perception

For many years a generally consistent conclusion reached by many authors in the field of psychology and sociology has been that social stereotyping is a totally undesirable strategy in making evaluations or behavioral predictions about persons or groups; the time and effort that may be saved is almost always at the expense of accuracy (e.g., Bogardus, 1950; Hayakawa, 1950). On the other hand, researchers in the field of person perception (Cline, 1964; Cronbach, 1955) have often reported that a large portion of the accuracy achieved by the judges in these studies can often be attributed to the judges' having an accurate stereotype on which they based their predictions.

Since the effect of global stereotyping on person perception accuracy has never been directly assessed, and since the findings of the person perception studies with regard to use of stereotypes suggest that the early negative conclusions reached about stereotyping and accuracy may not be entirely valid, the present study was conducted. An attempt was made to determine the relationship between accuracy in person perception and degrees of global stereotyping. An attempt was also made to assess some personality correlates of stereotyping since it was believed this knowledge would facilitate the understanding of the relationship between stereotyping and accuracy.

The following three hypotheses were advanced: (1) persons who

stereotype to a greater than average degree will tend to have lower judging accuracy scores than persons who stereotype at an average or below average level; (2)persons who stereotype at an above average level will differ from persons who stereotype at an average level on several personality traits, as measured by the California Psychological Inventory; (3) persons who stereotype at a below average level will tend to differ from persons who stereotype at an average level on several personality traits.

Ninety-seven male and female University of Utah student volunteers were asked to complete the CPI, view three of the interview films developed by Cline (1955) and complete the three corresponding judging accuracy tests, and complete two stereotyping tests that were developed for the purposes of this study. The original group of 97 students was divided into three subgroups: HS group (High Stereotypers--scores above average), AS group (Average Stereotypers--average scores), and LS group (Low Stereotypers--below average scores). The first hypothesis was tested by comparing the HS group mean judging accuracy score with the AS and LS group mean judging accuracy scores by means of one-tailed t-tests. The second and third hypotheses were tested conjointly by means of separate one-way analyses of variance of the three groups' scores on each of the CPI subscales.

The results of these statistical tests suggested that the first hypothesis should not be accepted. Of 24 comparisons of means, 22 yielded non-significant t-scores. The second and third hypotheses

were only minimally supported. The HS female group had a higher mean score on the Sociability subscale and the LS male group had higher mean scores on the Communality and Intellectual Efficiency subscales.

Correlating the stereotyping and judging accuracy scores led to the conclusion that the two variables are positively related. It would appear that in situations where judgments about others must be made on very limited data, use of stereotypes may be the most efficient strategy. It was also concluded that inferred or hypothesized relationships between stereotyping and constructs such as intellectual rigidity and cognitive simplicity should be re-examined.

We have presented here only a summary of this research and its findings. A completely detailed version is available as a Master's thesis from the University of Utah library under the name of Dale L. Penprase, datad June 1972.